

 **INSTITUTE OF
EDUCATION SCIENCES**

**NATIONAL CENTER FOR
EDUCATION STATISTICS**

Help



> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data

**School
Survey
On
Crime &
Safety**

SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data

 [Download Transcript](#)  [Download Slides](#)  [Download Glossary](#)

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Module Objectives

00:01:06

Module Objectives

- Describe the sampling procedures for SSOCS
- Describe SSOCS weights that must be applied to ensure estimates made from the data are representative of the study population
- Describe appropriate procedures for calculating standard errors
- Describe how imputation procedures were applied to missing data

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS

Help

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Sample Design: Creating the Sample Frame

00:00:55

SSOCS Sample Design: Creating the Sample Frame


- Defines the population from which a sample is drawn for research purposes
- SSOCS sampling frame is based on the most recent [Common Core of Data \(CCD\)](#)
 - For the 2010 SSOCS, this was the 2007-2008 CCD

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > SD: Creating the Sample Frame (Continued)

00:00:45

SSOCS Sample Design: Creating the Sample Frame (Continued)

- Certain types of schools from the CCD are excluded from the SSOCS sampling frame including:
 - Schools located outside of the 50 states and the District of Columbia
 - Bureau of Indian Education schools
 - Special education schools
 - Vocational schools
 - Alternative schools
 - Ungraded schools
 - Schools with a high grade of kindergarten or lower
- Regular schools, charter schools, and schools that have partial or total magnet programs are included in the frame

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS

Help

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Sample Design: Design Objectives

00:00:35

Sample Design: Design Objectives

- The same general sample design was used to select a sample of schools for SSOCS in 2000, 2004, 2006, 2008 and 2010
- The objective of the 2009–2010 sample design was twofold:
 - to obtain overall cross-sectional and subgroup estimates of important indicators of school crime and safety
 - to develop precise estimates of change in various characteristics relating to crime between the SSOCS administrations

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Sample Design: Stratification

00:00:34

Sample Design: Stratification

- To attain these objectives, researchers drew a stratified sample of regular public schools in each year of SSOCS
- Stratification is used to ensure that different subgroups are adequately represented in the sample
- Stratification involves dividing the sampling frame into relevant subgroups, or “strata”, prior to sample selection
- Stratification helps to increase accuracy when estimating population parameters for these subgroups by ensuring that different subgroups of a population are represented adequately in the sample

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Sample Design: Defining the Strata

00:00:43

Sample Design: Defining the Strata

- SSOCS strata were defined by school level, locale, and enrollment size
 - School level indicates whether the school enrolls students in primary, middle or high school grades, or some combination
 - Locale – characterized as City, Suburb, Town, or Rural – is based on the Census-defined geographic area in which the school is located
 - Enrollment size is categorized into four ranges based on the number of students attending
- These three explicit stratification variables have been shown to be related to school crime and thus create meaningful subgroups for this survey

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Defining the Strata (Continued)

00:00:56

Sample Design: Defining the Strata (Continued)

- In addition, region and percent White enrollment were used as implicit stratification variables by sorting schools by these variables within each explicit stratum before sample selection
 - region is the Census region of the school, categorized as the Northeast, South, Midwest, and West
 - percent White enrollment represents the percentage of students enrolled in the school who are identified as White, non-Hispanic
- Note that prior to the 2008-2009 data collection, the locale variable used to stratify the sample was derived from an 8-level metro based code
- In 2008, the variable was switched to one derived from a 12-level place based code consolidated into four categories: city, suburb, town and rural

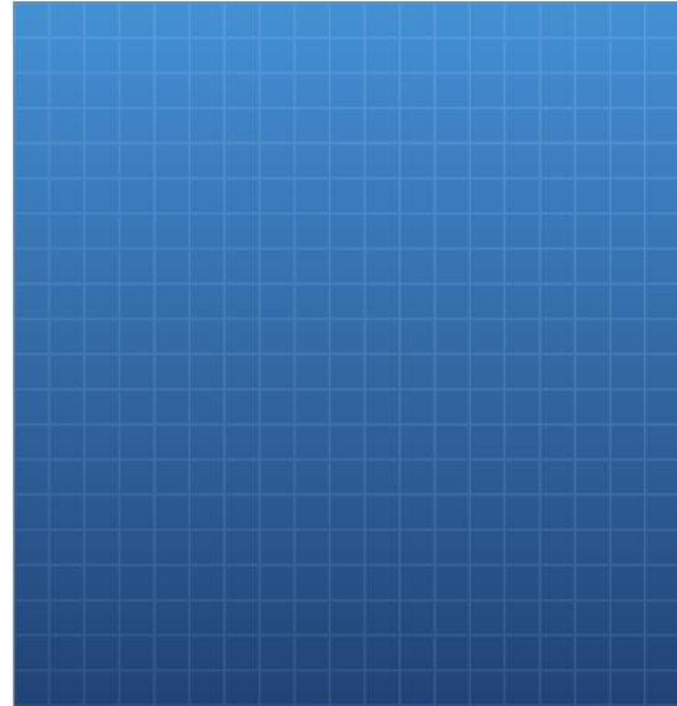
> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Sample Design: Drawing the Sample

00:02:07

SSOCS Sample Design: Drawing the Sample

Sample Design Process

- To determine how many schools to sample, NCES researchers define an optimal goal for the number of completed surveys to collect in order to meet their objectives
- Potential non-response must be taken into account
- Initial sample selected must be greater than the final goal



> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > SSOCS Sample Design: Sample Weights

00:01:42

SSOCS Sample Design: Sample Weights

- Weighting of the data is necessary to obtain population-based estimates, minimize non-response bias, and reduce sampling error
 - Sample weights allow for inferences to be made about the total population of schools from which the sampled schools are drawn
 - The method for calculating weights has not varied from year to year




> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Unit Response Rate and Weighting

00:02:40

Unit Response Rate and Weighting

- After the data are collected, a unit response rate is calculated for SSOCS
- The unit response rate is the ratio of completed eligible respondents to the total count of eligible respondents
- For SSOCS data, there are three measures to evaluate response:

INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS

Help

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Response Rate/Weighting (Continued)

00:01:28

Unit Response Rate and Weighting (Continued)Responding schools were compared to [non-responding schools](#) by:


> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Item Non-response

00:02:00

Item Non-response

SSOCS data set variables included in SSOCS 2010 Item Nonresponse Bias Analysis:

- C0326 Number of physical attacks or fights with a weapon (84.01 percent responded)
- C0330 Number of physical attacks or fights without a weapon (83.30 percent responded)
- C0332 Number of physical attacks or fights without a weapon reported to police (81.15 percent responded)

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS

Help

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Replication Techniques


00:01:46

Standard Error Calculation in SSOCS - Replication Techniques

- This method calculates appropriate SEs based on differences between estimates from the full sample and a series of created subsamples (replicates)
- Select replicate weights that are associated with the final sampling weight
- SSOCS replication weights use the Jackknife replication method
 - 2010 replicate weights are labeled 'REPWGT1' through 'REPWGT50'

[Standard errors](#) for selected SSOCS variables are calculated and reported on the NCES website for your reference



 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS

Help

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Taylor Series Linearization

00:00:45

Standard Error Calculation in SSOCS - Taylor Series Linearization

- This method uses primary sampling unit (PSU) and strata identifiers to calculate appropriate SEs
- Select the PSU and stratum variables (SCHID and STRATA) associated with the final sampling weight variable

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Missing Data and Editing

00:01:20

Missing Data and Editing



- Files containing missing data can be problematic for research
 - Analysis of incomplete datasets may cause users to arrive at different conclusions
 - Certain groups of respondents may be more likely than others to leave some survey items unanswered, creating bias in the survey estimates
- Data editing included:
 - Computer program-based editing and data retrieval follow-up interviewing were used in SSOCS to check for completeness and valid data ranges, consistency, and skip patterns
 - Data retrieval follow-up interviews were conducted for respondents missing critical survey items


> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Missing Data and Imputation

00:01:14

Missing Data and Imputation

Even after computer program-based editing and data retrieval interviewing, completed SSOCS surveys still contain some level of item nonresponse, accordingly, imputation procedures are used to create values for all questionnaire items' to the existing first statement on the slide

- The base-weighted item response rates for SSOCS 2010 were generally high
 - The mean weighted item response rate was about 98 percent
 - After editing, 99 percent of items in the 2010 data file had weighted response rates of over 85 percent
 - Most of the data in the SSOCS file are the originally entered values from the respondents
- Imputation methods were tailored to the nature of each survey item using either aggregate proportions, best match, or clerical methods
- Imputation flags are provided for each variable where imputation was used

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS

Help

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Imputation Methods - Aggregate Proportions

00:01:32

Imputation Methods - Aggregate Proportions

- Values estimated using data from respondents matched on key characteristics to the respondent with the missing data
- Rather than imputing counts from a single donor or a mean count from a group of donors, proportions were imputed using two methods

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Imputation Methods - Best Match

00:01:01

Imputation Methods - Best Match

School Characteristics: 2009-10 School Year	
25. As of October 1, 2009, what was your school's total enrollment?	1500 Students
26. What percentage of your current students fit the following criteria? If none, please place an "X" in the None box.	Percent of students
a. Eligible for free or reduced-price lunch	24 %
b. Limited English Proficient (LEP)	10 %
c. Special education students*	3 %
d. Male	48 %
27. What is your best estimate of the percentage of your current students who meet the following criteria? If none, please place an "X" in the None box.	Percent of students
a. Below the 15 th percentile on standardized tests	10 %
b. Likely to go to college after high school	56 %
c. Consider academic achievement to be very important	75 %
28. How many classroom changes do most students make in a typical day? Count going to lunch and then returning to the same or a different classroom as two classroom changes. Do not count morning arrival or afternoon departure. If none, please place an "X" in the None box.	4 Typical number of classroom changes

*Please use the definition on page 2.

School Characteristics: 2009-10 School Year	
25. As of October 1, 2009, what was your school's total enrollment?	1450 Students
26. What percentage of your current students fit the following criteria? If none, please place an "X" in the None box.	Percent of students
a. Eligible for free or reduced-price lunch	24 %
b. Limited English Proficient (LEP)	11 %
c. Special education students*	3 %
d. Male	48 %
27. What is your best estimate of the percentage of your current students who meet the following criteria? If none, please place an "X" in the None box.	Percent of students
a. Below the 15 th percentile on standardized tests	%
b. Likely to go to college after high school	56 %
c. Consider academic achievement to be very important	74 %
28. How many classroom changes do most students make in a typical day? Count going to lunch and then returning to the same or a different classroom as two classroom changes. Do not count morning arrival or afternoon departure. If none, please place an "X" in the None box.	4 Typical number of classroom changes

*Please use the definition on page 2.

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Imputation Methods - Clerical

00:00:31

Imputation Methods - Clerical

In some instances, missing data are available from other sources:

- CCD sampling frame used to impute values for those schools missing student enrollment data for item 25
- CCD data were also available on school type for item 31 and the percentage of students eligible for free or reduced-price lunch, item 26a
- In other instances, research was done on school administrative records to estimate logical values for missing data


The screenshot displays a portion of the CCD Common Core of Data survey form. At the top, the CCD logo is visible. Item 25 asks for the school's total enrollment as of October 1, 2009, with a response field showing '522' and the unit 'Students'. Item 26 asks for the percentage of current students fitting specific criteria, with a sub-item 26a for 'Eligible for free or reduced-price lunch' and a response field showing 'None'. Item 31 asks for the best description of the school, with a 'Check one response' instruction and five radio button options: 'Regular public school', 'Charter school', 'Has a magnet program for part of the school', 'Exclusively a magnet school', and 'Other - Please specify'. A response field for 'Other' is also present.

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Imputation Methods - Order of Use

00:00:42

Imputation Methods - Order of Use

- Interrelationships between the SSOCS survey items require that a specific imputation order be followed
 - For example, because item 25 – student enrollment – is used in imputation for other variables, this item is imputed first
- In some cases, values for a particular variable are limited by the values of other variables
 - For example, the matrix of disciplinary actions by offenses in item 23 is related to the total number of offenses recorded and the total number of disciplinary actions recorded in earlier questions
 - If values must be imputed for these earlier items, this is done before item 23

INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS

Help

> SSOCS Sample Design, Weights, Variance, and Imputed and Missing Data > Summary and Resources

00:00:50

Summary and Resources

This module provided information about the:

- SSOCS sampling procedures
- SSOCS weights that must be applied to ensure estimates made from the data are representative of the study population
- Appropriate procedures for calculating standard errors
- How imputation procedures were applied to missing data

Resources

- [Common Core of Data \(CCD\)](#)
- [Non-responding schools \(User's Manual\)](#)
- [Standard Errors](#)